

An Alternative Analysis for Crossover Studies that Accounts for Between-Group Disparities in Drug Response

Ton J. M. Cleophas

Departments of Medicine and Clinical Pharmacology, Merwede Hospital Dordrecht, Sliedrecht, The Netherlands

Summary: Background: In crossover clinical trials comparing completely different treatments, patients tend to fall into different populations: those who respond better to treatment 1 and those who respond better to treatment 2. The correlation between treatment response in such trials is negative. The current ANCOVA analysis for crossover studies does not allow for correlations being negative, and is therefore not adequate for testing in this kind of trials.

Objective of study: To study whether matrix algebra provides a more appropriate approach for this purpose.

Results and conclusions: Using a mathematical model as well as hypothesized examples, it is demonstrated that matrix algebra of 2 pairs of cells of the same order not only allows for negative correlations in a crossover design but also provides enough power to test both the treatment and carryover effect.

Introduction

Following Grizzle and Brown (1, 2), many workers (3–6) currently assume, for the analysis of crossover studies, the following statistical model based on analysis of covariance (ANCOVA):

$$y_{ijk} = \mu + \xi_{ij} + \pi_k + \phi_1 + \lambda_1 + \varepsilon_{ijk}$$

where: y_{ijk} is the response for the j^{th} subject in the i^{th} sequence during the k^{th} period

μ is the mean response

ξ_{ij} is the effect of the j^{th} subject in the i^{th} sequence

π_k is the effect of the k^{th} period

ϕ_1 is the direct effect of the 1st treatment

λ_1 is the carryover effect of the 1st treatment

ε_{ijk} is the independent error term

$$\text{cov}(y_{ijk}, y_{ijk}') = \begin{cases} \sigma_{\xi}^2 + \sigma_{\varepsilon}^2 & \text{if } k = k' \\ \sigma_{\xi}^2 & \text{if } k \neq k' \end{cases}$$

and derive:

$$\text{corr}(y_{ij1}, y_{ij2}) = \sigma_{\xi}^2 / (\sigma_{\xi}^2 + \sigma_{\varepsilon}^2) = \rho$$

The merit of this ANCOVA approach is that it provides a fairly straightforward way of carrying out an after-the-fact statistical adjustment of the data, to equate concomitant variables or covariates. However, the problems of this ANCOVA approach are:

1. it does not allow for correlations being negative, because $\sigma_{\xi}^2 / (\sigma_{\xi}^2 + \sigma_{\varepsilon}^2)$ simply cannot be negative,
2. it fails to include a covariate related to differences in the populations under study.

On the one hand, in a treatment comparison the new treatment may be a slight modification of the standard

or be equivalent to it with the addition of a new component. In this situation there is a positive correlation between the new and the standard treatment, i. e. treatment 1 will perform well when treatment 2 does so too. On the other hand, in trials with completely different treatments patients tend to fall into different populations: those who respond to treatment 1 and those who respond to treatment 2. Patients with angina pectoris unresponsive to beta-blockers generally respond either to calcium channel blockers or to nitroglycerines. Also hypertension, Raynaud's phenomenon, different types of cardiac arrhythmias and chronic obstructive pulmonary disease are known to be conditions where a non-response to a particular compound is frequently associated with an excellent response to a completely different compound. These are examples of situations in which a crossover study is likely to give rise to a strong negative correlation. This may even be so in spite of the general likelihood in any crossover of a somewhat positive correlation, because one subject is used for comparison of two treatments. The mechanism of between-group disparities has been well-recognized in clinical pharmacology; in fact, this is the main reason that in treatment protocols the principle of stepped care is currently being replaced by individualized care (7). The recognition of between-group disparities in drug response also implies that any mathematical approach for crossover studies should allow for correlations being negative, and that the above ANCOVA approach is not appropriate for that purpose. A more appropriate approach would be the application of basic matrix algebra in which 2×2 cells of the same order can be added or subtracted in a cell by cell manner. The current study investigates whether this ap-

proach, unlike the ANCOVA, allows for correlations being negative, and provides adequate power to test both treatment and carryover effects.

Statistical Model

According to *Scheffé* (8) the notion for a simple crossover study is:

	Period 1		Period 2	
	Treat- ment	Mean efficacy	Treat- ment	Mean efficacy
Group 1 (n_1)	1	$y_{1.1}$	2	$y_{1.2}$
Group 2 (n_2)	2	$y_{2.1}$	1	$y_{2.2}$

y_{ijk} = the response of the j^{th} patient in the i^{th} group in the k^{th} period. We assume that $n_1 = n_2 = n$.

$$y_{i.k} = \sum y_{ijk}/n$$

Also we assume:

1. that the samples have a normal distribution or a t-distribution,
2. that in a crossover without a carryover effect the data of the second period are a true reflection of the first period because the two treatment groups are symmetrical.

To test the treatment effect (ϕ) the sum of the results of treatment 1 is compared with the treatment 2 results ($y_{1.1} + y_{2.2}$ vs $y_{1.2} + y_{2.1}$). To trace the carryover effect (λ) the sum of the results in group 1 is compared with the group 2 results ($y_{1.1} + y_{1.2}$ vs $y_{2.1} + y_{2.2}$). To trace the time effect (π) the sum of the results in period 1 is compared with the period 2 results ($y_{1.1} + y_{2.1}$ vs $y_{1.2} + y_{2.2}$).

The null-hypotheses that ϕ , λ , and π are zero

$$\phi \quad [(y_{1.1} + y_{2.2}) - (y_{1.2} + y_{2.1})] = 0$$

$$\lambda \quad [(y_{2.1} + y_{2.2}) - (y_{1.1} + y_{1.2})] = 0$$

$$\pi \quad [(y_{1.1} + y_{2.1}) - (y_{1.2} + y_{2.2})] = 0$$

should be slightly remodelled into

$$\phi \quad [(y_{1.1} - y_{1.2}) - (y_{2.1} - y_{2.2})] = 0$$

$$\lambda \quad [(y_{2.1} + y_{2.2}) - (y_{1.1} + y_{1.2})] = 0$$

$$\pi \quad [(y_{1.1} - y_{1.2}) + (y_{2.1} - y_{2.2})] = 0$$

In this way 2×2 paired cells can be adequately added or subtracted in a cell by cell manner.

These null hypotheses can be tested, for example, by t-statistic or one-way analysis of variance (ANOVA). The larger the extent to which the t or F value of our distribution differs from zero, the more sensitivity and power the statistical approach provides.

$$t = \frac{d}{SE} \quad (\text{or one-way ANOVA, F value})$$

where d is ϕ , λ , or π , and SE is their standard error.

SE is calculated by use of the standard formulas for the variance (σ^2) of paired and unpaired sums and differences.

$$\sigma_{\text{paired sums}}^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$$

$$\sigma_{\text{paired differences}}^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

$$\sigma_{\text{unpaired sums}}^2 = \sigma_1^2 + \sigma_2^2$$

$$\sigma_{\text{unpaired differences}}^2 = \sigma_1^2 + \sigma_2^2$$

We assume that $\sigma = \sigma_{Y1.1} = \sigma_{Y1.2} = \sigma_{Y2.1} = \sigma_{Y2.2}$ = standard deviation of the samples in each of the cells, and that $\rho = \rho_{Y1.1 \text{ vs } Y1.2} = \rho_{Y2.1 \text{ vs } Y2.2}$ = correlation coefficient between the samples of each of the two paired cells.

Then

$$\sigma_{\phi}^2 = 2(2\sigma^2)(1 - \rho)$$

$$\sigma_{\lambda}^2 = 2(2\sigma^2)(1 + \rho)$$

$$\sigma_{\pi}^2 = 2(2\sigma^2)(1 - \rho)$$

Because $n_1 = n_2 = n$ we now can calculate the standard errors (SE) as follows:

$$\begin{aligned} SE_{\phi} &= \sqrt{4\sigma^2(1 - \rho)(1/2n + 1/2n)} \\ &= \sqrt{4\sigma^2(1 - \rho)/n} \end{aligned}$$

and accordingly

$$SE_{\lambda} = \sqrt{4\sigma^2(1 + \rho)/n}$$

$$SE_{\pi} = \sqrt{4\sigma^2(1 - \rho)/n}$$

Suppose $\lambda = \phi$ and $\rho = 0$, then $t_{\lambda} = t_{\phi}$. In this situation the powers to test carryover and the treatment effect are equal.

$$\text{If } \lambda = \phi \text{ and } \rho > 0 \quad \text{then } t_{\lambda} < t_{\phi}$$

$$\text{If } \lambda = \phi \text{ and } \rho < 0 \quad \text{then } t_{\lambda} > t_{\phi}$$

Thus, the powers of testing are largely dependent on the correlation between treatment modalities ρ . Whenever $\rho > 0$ we will have much more power to test the treatment effect than carryover effect of similar size. We should add that in practice $\sigma_{Y1.2}$ may be somewhat larger than $\sigma_{Y1.1}$, because the larger the data the larger the variances. If, e.g., $\sigma_{Y1.2}$ is 10% larger than $\sigma_{Y1.1}$, ρ will change from > 0 to > 0.05 . Thus, in this situation the level of positive correlation required tends to rise.

If the situation allows the assumption that the reference treatment in the trial is inert, it will not usually cause a carryover effect, and consequently we will be able to account for the carryover effect only in the group that received the active treatment after the inert treatment. Assuming that treatment 1 is inert then we will have to account for the carryover effect in group 2 only, and the appropriate analysis for carryover effect in this situation therefore becomes

$$\lambda (y_{2.2} - y_{1.1}) = 0$$

Standard error (SE) is calculated by use of the formula for unpaired differences

$$\begin{aligned}\sigma_{\lambda}^2 &= 2\sigma^2 \\ SE_{\lambda} &= \sqrt{2\sigma^2 (1/n + 1/n)} \\ SE_{\lambda} &= \sqrt{4\sigma^2/n}\end{aligned}$$

Again, if $\lambda = \varphi$ and $\rho = 0$ then $t_{\lambda} = t_{\varphi}$
 if $\lambda = \varphi$ and $\rho > 0$ then $t_{\lambda} < t_{\varphi}$
 if $\lambda = \varphi$ and $\rho < 0$ then $t_{\lambda} > t_{\varphi}$.

The time effect (π) is considered to influence the data of the two treatments in a similar way, and its influence on treatment differences is therefore negligible.

	Period 1		Period 2	
	Treat- ment	Mean response	Treat- ment	Mean response
Group 1	1	$y_{1.1}$	2	$y_{1.2} + \frac{1}{2}\pi$
Group 2	2	$y_{2.1}$	1	$y_{2.2} + \frac{1}{2}\pi$

Under the assumption $\lambda = 0$ we have

$$\begin{aligned}\varphi &= (y_{1.1} - y_{1.2} - \frac{1}{2}\pi) - (y_{2.1} - y_{2.2} - \frac{1}{2}\pi) \\ &= y_{1.1} - y_{1.2} - y_{2.1} + y_{2.2}\end{aligned}$$

While the estimate of the time or period effect may be relevant to clinicians its size apparently does not influence the analysis of the treatment data, and so it does not have to be taken into account.

Hypothetical Examples, Power Analysis

The power of testing the treatment effect is therefore heavily dependent on the variances of the paired differences. This can be illustrated by the example in figure 1. It shows how levels of correlation between two treatment modalities influence the variances of paired differences in opposite ways. In the case of a negative correlation the standard error of the paired differences largely outnumbers the standard error of the positive correlation situation, while the zero-correlation-standard error is in-

between. This is, of course, the main reason that finding a significant difference between two samples with a positive correlation is much easier than when the correlation is negative. Suppose we have 3 crossovers with the data of figure 1 in Group 1, and that Group 2 is a true reflection of Group 1.

	Period 1		Period 2	
	Treat- ment	Mean response	Treat- ment	Mean response
Group 1 n = 10	Vaso- dilator 1	$y_{1.1} 35 \pm 10$	Vaso- dilator 2	$y_{1.2} 25 \pm 10$
Group 2 n = 10	Vaso- dilator 2	$y_{2.1} 25 \pm 10$	Vaso- dilator 1	$y_{2.2} 35 \pm 10$

In the given example improvement is measured by the number of *Raynaud* attacks/week. Vasodilator 2 is more efficient than vasodilator 1, because there are fewer attacks. Suppose, this result causes a physical carryover effect in Group 2 from period 1 into period 2. Then, the mean value at $y_{2.2}$ changes from 35 into $35 - \lambda$, where λ = mean value of carryover effect in Group 2 of the trial. If, for the purpose of this particular example, we assume that the variance of λ is zero, it can be simply subtracted from the paired differences of the groups without influencing their standard errors. It does not invalidate the overall conclusions of the procedure, because any variance larger than zero has to be added to the variances already in the study, thus further reducing the power of testing. This assumption therefore produces the largest powers for the given data. Calculations are carried out by paired and unpaired t-statistics, and power is approximated from the equation

$$\text{POWER} = 1 - \beta = 1 - \text{prob} [Z \leq (t^1 - t)]$$

where Z represents the standardized value for the differences between two means, t^1 represents the upper critical value of t for the given degrees of freedom, and α has been specified (treatment effect $\alpha = 0.05$, carryover effect $\alpha = 0.10$) according to *Grizzle* (1). In figure 2

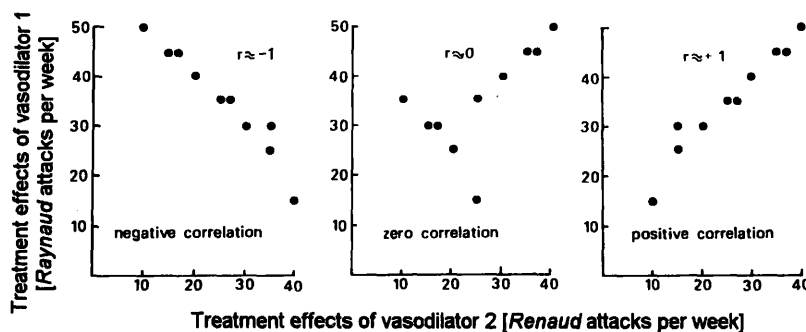


Fig. 1 Individual values of paired observations with similar group means, n's and SD's (vasodilator 1: 35 ± 10 , vasodilator 2: 25 ± 10).

negative correlation ($\rho \approx -1$): SE paired differences 6.8
 zero correlation ($\rho \approx 0$): SE paired differences 2.9
 positive correlation ($\rho \approx +1$): SE paired differences 0.8

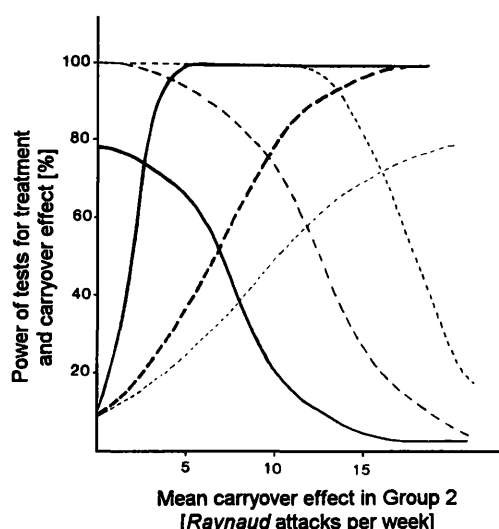


Fig. 2 Power graphs of tests for treatment effect (slope downwards) and carryover effect (slope upwards) of the three crossover situations from figure 1 with the variable λ (mean carryover effect in group 2) added.

— $\rho \cong -1$; - - $\rho \cong 0$; $\rho \cong +1$

power graphs are drawn of tests for treatment and carryover effects of the three crossover situations from figure 1 with the variable λ added. First, there are three power curves of the treatment effect for the three correlation situations. As λ increases, all three gradually come down. The negative correlation curve is the first to do so. Consequently, this situation has generally little power of coming to the correct conclusion. At $\lambda = 10$, when the treatment effect is equal to the carryover effect, there is less than 30% power left. This means that there is a more than 70% chance that the treatment effect is erroneously unobserved in this study. Considering that a power of approximately 80% is required for reliable testing, we cannot test carryover here in a sensitive manner. The zero and positive correlation situations provide essentially more power.

There are also three power curves of carryover effect for the three correlation situations. The negative correlation curve provides essentially better power than the zero and positive correlation curves do. This example shows that strong positive correlations leave little power to test a carryover effect. It also shows that strong negative correlations produce excessive power to test a carryover effect.

Figure 3 gives the alternative approach where a carryover effect is tested according to the one-group carryover principle. In this approach the three different correlation situations have one and the same carryover curve because the unpaired comparison for carryover has identical means, standard deviations, and n 's, and is therefore not influenced by different levels of correlation. It is actually the same curve as the $\lambda = 0$ curve from figure 2. Its power increases with λ . At $\lambda = 10$, when the

treatment effect is equal to the carryover effect, it is about 80%.

The latter approach should be performed instead of the former whenever it is permitted by the modalities of treatment. This is because it both offers less power with negative correlations and more power with positive correlations. In so doing it prevents the detection of small and clinically irrelevant carryover effects, and at the same time the detection of small and clinically unimportant treatment effects.

Discussion

Crossover studies are routinely used in clinical research. For example no less than 22% of the double-blind placebo-controlled hypertension trials published in 1993 were crossover studies (9, 10). A major advantage of the crossover design is that it eliminates between-subject variability of symptoms. However, problems include the occurrence of carryover and time effects, and the fact that the design itself offers little power to test such effects. Power of testing such effects is enhanced by the use of ANCOVA which is able to test treatment, carryover and time effects simultaneously (3–6). ANCOVA in its current form does not allow, however, for correlations being negative which is an important possibility in studies comparing completely different treatments. Adding another covariate to adjust this flaw weakens the ANCOVA and makes it hardly testable. An alternative approach is the classic two-stage analysis in which carryover and time effects are tested first and the treatment effect second. The power of this approach is enhanced when data are analysed by use of matrix algebra. Data

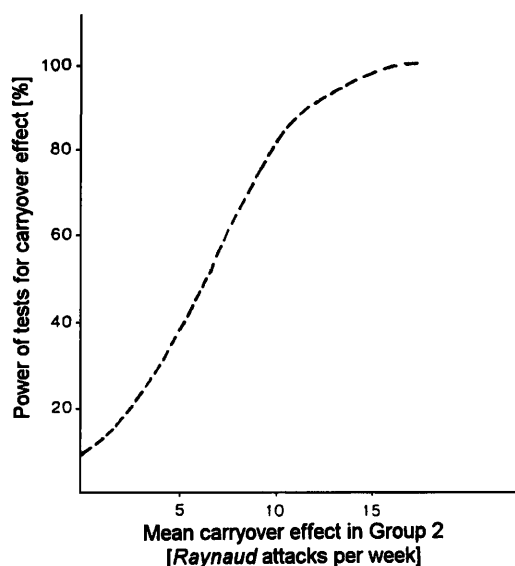


Fig. 3 Power graph of tests for carryover effect of the three crossover situations from figure 1 with the variable λ (mean carryover effect in group 2) added.

The graph displays a carryover effect irrespective of the size of ρ (one-group carryover effect analysis).

are analysed in the form of 2×2 cells of the same order, which are added or subtracted in a cell by cell manner.

The alternative analysis for crossover studies as presented in this paper includes the following steps:

1. Carryover and treatment effects in a crossover study are tested by use of matrix algebra, in which 2 pairs of cells of the same order are added or subtracted.
2. A one-group analysis for the carryover effect is performed whenever the clinical situation permits it. It generally offers a more adequate power for testing the data.
3. While the estimate of time effects may be relevant to clinicians its size does not influence the estimate of the treatment effects, and, therefore, does not have to be taken into account in the analysis of the treatment data.
4. In crossover trials with a zero or positive correlation, small carryover effects hardly reduce the power of demonstrating a treatment effect. Nevertheless, it makes

sense to test for a carryover effect in the group that received the effective compound first. If such testing is positive, a parallel-group analysis of period 1 of the trial can effectively be used for demonstrating a treatment effect.

In the event of a significant carryover effect in group 2, a parallel-group analysis of period 1 will provide a significant test as well, and it will do so at the same or even at a higher level of significance than the test for a carryover effect. This is because when the carryover effect is maximal, the mean response in period 2 group 2, is that $y_{2.2}$ equals $y_{2.1}$. The between-group difference between $y_{2.1}$ and $y_{1.1}$ in period 1 will, therefore, be at least as large as the difference between $y_{2.2}$ and $y_{1.1}$ but probably larger. Therefore, no further test for a treatment effect seems to be required in this situation, and the data of period 2 could be disregarded. If we still wish to enhance the power of testing this situation, we could perfectly well make use of the unbiased period-2-data of group 1 using a weighting procedure (11).

References

1. Grizzle JE. The two-period change-over design and its use in clinical trials. *Biometrics* 1965; 22:469–80.
2. Brown BW. The crossover experiment for clinical trials. *Biometrics* 1980; 36:69–79.
3. Grieve AP. A bayesian analysis of the two-period crossover design for clinical trials. *Biometrics* 1985; 41:979–90.
4. Willan AR, Pater JL. Carryover and the two-period crossover clinical trial. *Biometrics* 1986; 42:593–9.
5. Freeman PR. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Stat Med* 1989; 8:1421–32.
6. Fleiss JL. A critique of recent research on the two-treatment crossover design. *Control Clin Trials* 1989; 10:237–44.
7. Nies AS, Spielberg SP. Individualization of drug therapy. In: Hardman JL, Limbird LE, editors. *Goodman and Gilman's pharmacological basis of therapeutics*. New York: McGraw-Hill, 1996:43–63.
8. Scheffé H. Mixed models. In: Scheffé H, editor. *The analysis of variance*. New York: Wiley & Sons, 1959:261–91.
9. Cleophas TJ. Clinical trials: relevance of correlation between treatment responses. *Eur J Clin Pharmacol* 1996; 50:1–6.
10. Jackson PR, Yeo WW, Cleophas TJ. Crossover trials. *Br J Clin Pharmacol* 1996; 42:399–404.
11. Cleophas TJ. Crossover studies: a modified analysis with more power. *Clin Pharmacol Ther* 1993; 53:515–20.

Received April 21/July 10, 1997

Corresponding author: T. J. M. Cleophas, Department of Medicine, Merwede Hospital Dordrecht, P.O. Box 97, NL-3360 AB Slidrecht, The Netherlands

